FSMNLP 2011

Proceedings of the

9th International Workshop Finite State Methods and Natural Language Processing

July 12–15, 2011
Université François Rabelais Tours
Blois, France

Sponsors:

©2011 The Association for Computational Linguistics

Order copies of this and other ACL proceedings from:

Association for Computational Linguistics (ACL)
209 N. Eighth Street
Stroudsburg, PA 18360
USA
Tel: +1-570-476-8006
Fax: +1-570-476-0860
acl@aclweb.org

Preface

These proceedings contain the papers presented at the 9th International Workshop on Finite State Methods and Natural Language Processing (FSMNLP 2011), which was held in Blois (France), July 12–15, 2011, jointly with the 16th International Conference on Implementation and Application of Automata (CIAA 2011).

The workshop covers a wide range of topics from morphology to stringology to formal language theory. This volume contains the 14 regular and 3 short papers that were presented at the workshop. In total, 30 papers (25 regular and 5 short papers) were submitted to a doubly blind refereeing process, in which each paper was reviewed by 3 program committee members. The overall acceptance rate was 57%. The program committee was composed of internationally leading researchers and practitioners selected from academia, research labs, and companies.

The organizing committee would like to thank the program committee for their hard work, the referees for their valuable feedback, the invited speakers for their innovative contributions, and the local organizers for their tireless efforts. We are particularly grateful for significant sponsorship from the *Campus de la CCI de Loir-et-Cher*, the *Université François-Rabelais Tours*, the *Centre National de la Recherche Scientifique*, the *Région Centre*, the city of *Blois*, the *Université de Rouen*, the *Université Paris-Est Marne-la-Vallée*, the *Communauté d'Agglomération de Blois* (Agglopolys), the *Ministère de l'Enseignement Supérieur et de la Recherche*, *Humanis*, the *Université d'Orléans* and *MAIF*.

MATTHIEU CONSTANT
ANDREAS MALETTI
AGATA SAVARY

Organizers:

Jean-Yves Antoine, Université François Rabelais Tours (France)
Béatrice Bouchou-Markhoff, Université François Rabelais Tours (France)
Pascal Caron, Université de Rouen (France)
Jean-Marc Champarnaud, Université de Rouen (France)
Matthieu Constant, Université Paris-Est Marne-la-Vallée (France), FSMNLP chair
Nathalie Friburger, Université François Rabelais Tours (France)
Mirian Halfeld Ferrari Alves, Université d'Orléans (France)
Aurore Leroy, Université François Rabelais Tours (France)
Andreas Maletti, University of Stuttgart (Germany)
Patrick Marcel, Université François Rabelais Tours (France)
Denis Maurel, Université François Rabelais Tours (France)
Veronika Peralta, Université François Rabelais Tours (France)
Yacine Sam, Université François Rabelais Tours (France)
Agata Savary, Université François Rabelais Tours (France), CIAA chair

Invited Speakers and Tutorialists:

Eric Laporte, Université Paris-Est Marne-la-Vallée (France)
Sylvain Lombardy, Université Paris-Est Marne-la-Vallée (France)
Mark-Jan Nederhof, University of St Andrews (United Kingdom)
Joachim Niehren, INRIA Lille (France)
Sheng Yu, University of Western Ontario (Canada)

Program Committee:

Cyril Allauzen, Google Inc. (USA)
Francisco Casacuberta, Instituto Tecnológico De Informática (Spain)
David Chiang, ISI, University of Southern California (USA)
Maxime Crochemore, King's College London (United Kingdom)
Jan Daciuk, Gdańsk University of Technology (Poland)
Frank Drewes, Umeå University (Sweden)
Dafydd Gibbon, University of Bielefeld (Germany)
Thomas Hanneforth, University of Potsdam (Germany)
Colin de la Higuera, University of Nantes (France)
Jan Holub, Czech Technical University in Prague (Czech Republic)
André Kempe, CADEGE Technologies & Consulting (France)
András Kornai, Eötvös Loránd University (Hungary)
Derrick Kourie, University of Pretoria (South Africa)
Eric Laporte, Université Paris-Est Marne-la-Vallée (France)
Sylvain Lombardy, Université Paris-Est Marne-la-Vallée (France)
Andreas Maletti, University of Stuttgart (Germany)
Mike Maxwell, University of Maryland (USA)
Kemal Oflazer, Carnegie Mellon University (Qatar)
Jakub Piskorski, Polish Academy of Sciences, Warsaw (Poland)
Laurette Pretorius, University of South Africa (South Africa)
Strahil Ristov, Ruđer Bošković Institute, Zagreb (Croatia)
Jim Rogers, Earlham College, Richmond (USA)
Giorgio Satta, University of Padua (Italy)
Max Silberztein, Université de Franche-Comté (France)
Bruce Watson, Universities of Pretoria and Stellenbosch (South Africa)
Anssi Yli-Jyrä, University of Helsinki (Finland)
Sheng Yu, University of Western Ontario (Canada)
Menno van Zaanen, Tilburg University (Netherlands)
Lynette van Zijl, Stellenbosch University (South Africa)

Additional Reviewers:

Hasan Ibne Akram
Bernd Bohnet
Fabienne Braune
Loek Cleophas
Yuan Gao
Carlos Gómez-Rodríguez
Jan Janousek
Slim Mesfar
Ernest Ngassam
Petr Prochazka
Tinus Strauss

Table of Contents

Intersection for Weighted Formalisms
Mark-Jan Nederhof 1

Modularization of Regular Growth Automata
Christian Wurm 3

Finite-state Representations Embodying Temporal Relations
Tim Fernando 12

Supervised and Semi-Supervised Sequence Learning for Recognition of Requisite Part and Effectuation Part in Law Sentences
Le-Minh Nguyen, Ngo Xuan Bach and Akira Shimazu 21

Compiling Simple Context Restrictions with Nondeterministic Automata
Anssi Yli-Jyrä 30

Constraint Grammar Parsing with Left and Right Sequential Finite Transducers
Mans Hulden 39

E-Dictionaries and Finite-State Automata for the Recognition of Named Entities
Cvetana Krstev, Duško Vitas, Ivan Obradović and Miloš Utvić 48

A Practical Algorithm for Intersecting Weighted Context-free Grammars with Finite-State Automata
Thomas Hanneforth 57

Open Source WFST Tools for LVCSR Cascade Development
Josef R. Novak, Nobuaki Minematsu and Keikichi Hirose 65

Intersection of Multitape Transducers vs. Cascade of Binary Transducers: The Example of Egyptian Hieroglyphs Transliteration
François Barthélemy and Serge Rosmorduc 74

A Note on Sequential Rule-Based POS Tagging
Sylvain Schmitz 83

FTrace: A Tool for Finite-State Morphology
James Kilbury, Katina Bontcheva and Younes Samih 88

Incremental Construction of Millstream Configurations Using Graph Transformation
Suna Bensch, Frank Drewes, Helmut Jürgensen and Brink van der Merwe 93

Stochastic K-TSS Bi-Languages for Machine Translation
M. Inés Torres and Francisco Casacuberta 98

Measuring the Confusability of Pronunciations in Speech Recognition
Panagiota Karanasou, François Yvon and Lori Lamel 107

Fast Yet Rich Morphological Analysis
Mohamed Altantawy, Nizar Habash and Owen Rambow 116

An Open-Source Finite State Morphological Transducer for Modern Standard Arabic
Mohammed Attia, Pavel Pecina, Antonio Toral, Lamia Tounsi and Josef van Genabith 125

Recognition and Translation of Arabic Named Entities with NooJ Using a New Representation Model
Héla Fehri, Kais Haddar and Abdelmajid Ben Hamadou 134

Conference Program

Tuesday, July 12, 2011

9:00–9:30 Opening

9:30–10:30 *Intersection for Weighted Formalisms*
Mark-Jan Nederhof

11:00–12:00 Tutorial by Sylvain Lombardy

14:00–14:30 *Modularization of Regular Growth Automata*
Christian Wurm

14:30–15:00 *Finite-state Representations Embodying Temporal Relations*
Tim Fernando

15:00–15:30 *Supervised and Semi-Supervised Sequence Learning for Recognition of Requisite Part and Effectuation Part in Law Sentences*
Le-Minh Nguyen, Ngo Xuan Bach and Akira Shimazu

16:00–16:30 *Compiling Simple Context Restrictions with Nondeterministic Automata*
Anssi Yli-Jyrä

16:30–17:00 *Constraint Grammar Parsing with Left and Right Sequential Finite Transducers*
Mans Hulden

17:00–17:30 *E-Dictionaries and Finite-State Automata for the Recognition of Named Entities*
Cvetana Krstev, Duško Vitas, Ivan Obradović and Miloš Utvić

Wednesday, July 13, 2011

9:30–10:30 Invited Talk by Joachim Niehren

11:00–12:00 Tutorial by Sylvain Lombardy

12:30–13:00 FSMNLP business meeting

14:30–15:00 *A Practical Algorithm for Intersecting Weighted Context-free Grammars with Finite-State Automata*
Thomas Hanneforth

15:00–15:30 *Open Source WFST Tools for LVCSR Cascade Development*
Josef R. Novak, Nobuaki Minematsu and Keikichi Hirose

15:30–16:00 *Intersection of Multitape Transducers vs. Cascade of Binary Transducers: The Example of Egyptian Hieroglyphs Transliteration*
François Barthélemy and Serge Rosmorduc

17:00–18:30 Guided Tour

Thursday, July 14, 2011

9:00–10:00 Invited Talk by Sheng Yu

10:00–11:00 Tutorial by Eric Laporte

11:30–12:30 Tutorial by Sylvain Lombardy

15:00–23:15 Excursion and Gala Dinner

Friday, July 15, 2011

9:00–9:20 *A Note on Sequential Rule-Based POS Tagging*
Sylvain Schmitz

9:20–9:40 *FTrace: A Tool for Finite-State Morphology*
James Kilbury, Katina Bontcheva and Younes Samih

9:40–10:00 *Incremental Construction of Millstream Configurations Using Graph Transformation*
Suna Bensch, Frank Drewes, Helmut Jürgensen and Brink van der Merwe

10:00–11:00 Tutorial by Eric Laporte

11:30–12:00 *Stochastic K-TSS Bi-Languages for Machine Translation*
M. Inés Torres and Francisco Casacuberta

12:00–12:30 *Measuring the Confusability of Pronunciations in Speech Recognition*
Panagiota Karanasou, François Yvon and Lori Lamel

14:30–15:00 *Fast Yet Rich Morphological Analysis*
Mohamed Altantawy, Nizar Habash and Owen Rambow

15:00–15:30 *An Open-Source Finite State Morphological Transducer for Modern Standard Arabic*
Mohammed Attia, Pavel Pecina, Antonio Toral, Lamia Tounsi and Josef van Genabith

15:30–16:00 *Recognition and Translation of Arabic Named Entities with NooJ Using a New Representation Model*
Héla Fehri, Kais Haddar and Abdelmajid Ben Hamadou

16:00–16:30 Closing

16:30–18:00 SIGFSM business meeting